

Kultura i Wartości

ISSN 2299-7806

Nr 41 (2026)

<http://dx.doi.org/10.17951/kw.2026.41.53-73>

An Enactive Account of Anger and Forgiveness

Daniel Shargel

Lawrence Technological University

e-mail: dshargel@ltu.edu



<https://orcid.org/0000-0002-6476-8938>

Anger, according to many theorists, represents an act of wrongdoing. Forgiveness puts an end to anger, and it is often earned by repairing the harm done to the victim. However, it is hard to square this view of anger with the following facts: that successfully repairing the harm is neither necessary nor sufficient for earning forgiveness, and that it is necessary for the wrongdoer to demonstrate their future trustworthiness. I will present an explanation of the relationship between anger and forgiveness based on an enactive theory of emotion. On this view, an emotion is fitting if it prepares the agent to effectively perform a relevant class of actions. Anger mobilizes the victim's mind and body to secure concessions from the wrongdoer. Forgiveness marks the moment when it is no longer fitting for the agent to prepare for aggressive action towards the wrongdoer.

Keywords: emotion, affordances, anger, forgiveness, moral psychology

1. Introduction

According to typical evaluative theories of emotions, anger represents wrongdoing. Forgiveness seems, at a minimum, to correlate with the cessation of anger. In many cases the wrongdoing produced harm, and that harm remains, or at least, a substantial portion of the harm remains, even after the wrongdoer has apologised and been forgiven. If anger represents wrongdoing, why should our anger end even if the wrongdoer has done harm, and the harm remains?

One solution is to build a forgiveness clause into the formal object of anger. On this view, anger is not simply directed at wrongdoing, it is rationally directed at a wrong that has not been forgiven. However, this move seems ad hoc. What is special about forgiveness, such that we should be angry before but not after? How does forgiveness relate to the prior wrongdoing?

There is a further puzzle concerning the forward-looking aspects of forgiveness. We earn forgiveness by repairing past damage, when possible, but repairing past damage is not only not necessary, it is also insufficient in many cases without an effort to gain forward-looking credibility. The efforts we make to repair past damage seem to be relevant for earning forgiveness to the extent that they demonstrate future trustworthiness, rather than depending on success at repairing the past harm.

It is also helpful to consider the perspective of the wrongdoer. We think that we have an obligation to seek forgiveness from those we have wronged, and that obligation seems to be grounded in the past, not the future. However, we satisfy this obligation by demonstrating that we will be trustworthy in the future.

I will propose an explanation of these observations that is based on adopting an enactive view of anger, and emotions more generally.¹ On this view, emotions are not in the business of representing features of the world. Instead, they are in the business of helping us change the world. The victim's anger helps them extract concessions from the wrongdoer. These concessions include repairing the past harm, within reason, and credibly demonstrating good intentions for the future. It is appropriate, even necessary, for the anger to end when the wrongdoer has made these concessions. This forgiveness provides the wrongdoer a number of benefits: the end of hostilities from the victim, the possibility of repairing their relationship with the victim, and the restoration of their moral stature.

The dynamics of anger and forgiveness are rooted in our shared cognitive and affective nature, and that is why it is possible to develop a broad theory about these phenomena. However, these dynamics are also shaped by our norms, including moral norms, prudential norms, and norms of proper etiquette, so in practice our requirements for extending forgiveness and our motivations for seeking forgiveness can vary quite widely. I might forgive, or seek forgiveness, because

¹ See Daniel Shargel and Jesse J. Prinz, "An Enactivist Theory of Emotional Content," in *The Ontology of Emotions*, ed. Hichem Naar and Fabrice Teroni (London: Routledge, 2018).

I want to save my job, or save my marriage, or save my soul. For most of the paper I will develop a single account of the dynamics that minimizes the moving parts. The goal of this account is to explain the puzzling features that I have described above. I will then describe some ways that additional factors can be added to the account to explain the diverse expressions of anger and forgiveness that we encounter in the wild.

2. Enactive Emotions

The basic idea behind an enactive theory of emotions is that each emotion type prepares us to effectively perform different types of actions.² I will focus on three aspects of emotion: bodily arousal, motivation, and social display, since they are sufficient for illustrating the basic anger-forgiveness dynamic, but other emotional features, such as valence and attentional direction, are also significant.

Anger is a high arousal state, and being in a high-arousal state 1) makes us more effective at acting vigorously, and 2) motivates us to act vigorously. To illustrate this, imagine doing a set of push-ups in a high arousal or a low arousal state. You would do more push-ups, and be more willing to do push-ups, in the former condition.

This integration of capability and motivation is beneficial. When we decide what to do, it is rational to consider the difficulty of performing an action, and the likelihood of success. If you are in a physiological state that temporarily decreases the difficulty and increases your likelihood of success at a certain class of actions, it is beneficial to enter a temporary motivational state that increases your willingness to perform any action that falls into that class.

Certain emotions are typically elicited when we need to prepare for vigorous actions. This includes anger—when we need to prepare to fight, fear—when we need to prepare to escape, and joy—when we have the opportunity to obtain benefits. Other emotions are typically elicited when we do not need to prepare for vigorous action: calm—when there is nothing that we are lacking, and sadness—when we have no opportunity to regain what we have lost. In these latter cases it is rational to de-mobilize the body. These physiological changes create or enhance

² See Shargel and Prinz, “An Enactivist Theory of Emotional Content.”

motoric affordances, since they temporarily change the ways that we are able to interact with our environment.³

Emotional expressions facilitate classes of actions as well. First, because these expressions are reliably integrated with other features of an emotion, they inform others about what we are both physically and motivationally prepared to do. In the case of anger, they send the message to whoever might interfere with us that we are ready, willing, and able to aggressively pursue our goals. This discourages others from interfering with us, since our anger makes us more difficult, expensive, and dangerous to interfere with. It also encourages others to collaborate with us if they share our goals, since these expressions indicate that we will be forceful and persistent.

We can provide similar explanations for other emotions. For example, expressions of joy attract friends and collaborators since they indicate that we are strong and magnanimous. This in turn amplifies our ability to make the most of opportunities. In sadness we are unable to help ourselves, and we lack the motivation to try, but we can express our sadness to attract help and care. In an enactive framework we can say that emotional expressions create or enhance social affordances, since they change the way we can interact with other people.⁴

3. Fittingness

The most widely accepted theories of emotion among philosophers take emotions to have a world-to-mind direction of fit.⁵ These views are diverse: some take fear to be a judgment, others take it to be a perception, while others take it to be a *sui generis* mental attitude, but they agree that fear is a representational state, and its success depends on its accuracy.

The enactive theory of emotions that I have presented above does not take emotions to have a world-to-mind direction of fit, and I will eventually argue that this is a strength for explaining forgiveness. However, before turning to the topic

³ See Shargel and Prinz, “An Enactivist Theory of Emotional Content.”

⁴ See *ibid.*

⁵ See Julien Deonna and Fabrice Teroni, *The Emotions: A Philosophical Introduction* (London: Routledge, 2012); John R. Searle, *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983).

of forgiveness I will clarify two issues about the success conditions of emotions on the enactive view. First, what are those conditions? Second, how can an enactive theory explain the compelling impression that emotions have a world-to-mind direction of fit?

The view is labeled ‘enactive’ because it holds that emotions bring something into being, rather than representing something.⁶ What they create are affordances—possibilities to act.⁷ In some cases they transform or enhance existing affordances, because the agent could perform the action without the emotion, but the action would be more difficult, less efficient, or less likely to be successful.

I’ll use an example to explore these issues further. Blake, a bully, shoves Avery. In response, Avery’s bodily arousal, integrated with their motivations, puts them in a position to fight back harder, with a set of expressions that make this transformation evident to Blake. Avery may or may not actually fight back physically against Blake. If Avery fights back, Avery’s anger may or may not allow Avery to overcome Blake in that struggle. Regardless of who wins, fighting back may or may not produce the best outcome for Avery from among the options that were available at the time. Given all of these possibilities, how should we define success for the emotion in an enactive framework?

It will be helpful here to adapt a concept from D’Arms and Jacobson⁸, who distinguish the fittingness of emotions from other norms such as morality or strategic benefit. It might not be moral for Avery to fight back against Blake, if they could instead, through compassion and wisdom, settle the conflict peacefully. It might not be strategic to fight back, because fighting exposes Avery to a variety of dangers. However, it might still be fitting for Avery to prepare to fight back, because Blake is a wrongful aggressor, fighting back is generally a relevant response to wrongful aggression, and visibly preparing to fight back is in many cases beneficial when facing a would-be aggressor.

⁶ See Francisco J. Varela, “Whence the Origin of Perception? A Cartography of Current Ideas,” in *Understanding Origins: Contemporary Ideas on the Origin of Life, Mind and Society*, ed. Francisco J. Varela and Jean-Pierre Dupuy (Boston: Kluwer, 1991).

⁷ See James J. Gibson, *The Ecological Approach to Visual Perception* (Boston: Houghton Mifflin, 1979).

⁸ See Justin D’Arms and Daniel Jacobson, “The Moralistic Fallacy: On the Appropriateness of Emotions,” *Philosophical and Phenomenological Research* 61, no. 1 (2000).

Typically we would analyze the fittingness of an emotion by identifying its formal object, and considering whether its particular object matches this formal object.⁹ Blake's aggression wronged Avery, so Avery's anger towards Blake is fitting. This presupposes that the emotion is a representational state with a world-to-mind direction of fit, so I will take a different approach.

In order to analyze the fittingness of enactive emotions, I will develop an analogy with modes on a camera. For starters, assume that a camera has a fixed number of modes.¹⁰ One setting is optimized for capturing landscapes, another for sports photography, another for portraits, another for close-ups of small objects, and another for low-light conditions. You carry around this camera all the time, and you don't know when you will want to take a photo. When an interesting shot presents itself you may or may not have time to switch modes, so you regularly adjust the mode depending on where you are and what photo opportunities you anticipate. If you are attending a basketball game you switch to sports mode, if you are taking a hike you switch to landscape mode, and if the family is assembled for a holiday dinner you switch to portrait mode.

Your preparations might fail in a number of different ways. Perhaps you are at a basketball game, so you put the camera in sports mode, but you actually want to take a selfie with your friends. Or you are on a hike, so you put the camera in landscape mode, and then need to quickly switch to close-up mode to take a picture of an interesting insect. In these cases the camera mode is not fitting.

In another case you are taking a photo in low light, and you are in low light mode. However, the photo is poor, because there was not enough light for the camera to function well, even in that mode. You might blame the camera for not having a good enough low light mode, or you might blame yourself for not realizing that the light was not sufficient, but there is no reason to criticize yourself

⁹ See Anthony Kenny, *Action, Emotion, and Will* (New York: Humanities Press, 1963).

¹⁰ In this example the camera has a discrete number of modes, which corresponds to a basic emotions theory. However, the example would work just as well with camera modes that vary on a number of dimensions, which would correspond with a dimensional theory of emotions, or with modes that are a function of a number of independent binary features, which would correspond to certain appraisal theories. See James A. Russell, "Core Affect and the Psychological Construction of Emotion," *Psychological Review* 110, no. 1 (2003); Klaus R. Scherer, "The Dynamic Architecture of Emotion: Evidence for the Component Process Model," *Cognition and Emotion* 23, no. 7 (2009).

for having the camera in that mode. It was the best mode available under the circumstances.

How should we describe the functions of these camera modes? There is a correspondence between camera modes and situation types, where the modes are fitting when they match, and unfitting when they do not. One might infer, based on these correspondences, that the modes represent situation types.

However, this is clearly incorrect. The function of the modes is to modify the mechanisms of the camera to take certain types of photos. The modes tend to co-occur with situation-types because those situations predict the opportunity to take types of photos. The photographer represents the situation types, and then causes the modes to correspond with the situation-types to maximize the likelihood of taking a successful photo, but the modes do not themselves represent anything.

Each emotion-type is analogous to a camera mode. Just like the modes of the camera have the function of preparing the camera to take certain types of pictures effectively, emotions have the function of preparing the agent to perform certain types of actions effectively. The camera modes are fitting when the camera is in the right mode for taking the types of photos that might be needed. Emotions are fitting when the agent is in the right type of emotional state for facilitating the types of actions they might need to perform.

Certain mental states of the photographer normally cause the camera to be in the right mode for the current situation. Those mental states are obviously distinct from the camera modes themselves. A different class of mental states (often called ‘appraisals’) normally cause the agent to be in the right emotional state for the current situation. Those appraisals are distinct from the emotions themselves, for the same reason that the camera modes are distinct from the mental states of the photographer that cause them.

In the camera example there is little temptation to theorize that the mental states that lead the photographer to select a camera mode are part of a unified state with that camera mode. One is a mental state, realized in the photographer’s brain, and the other is a camera state, realized in the camera, so assimilating the two would only appeal to a truly committed extended mind theorist. However, the appraisals that produce emotional states, and the suite of psychological and physiological changes that occur during emotional states, are all features of the same individual. Emotions already seem to be a messy combination of dissimilar features, so it is not obvious why some should be included, and others should be

excluded. Furthermore, it is very appealing to include the appraisal as a constituent of the emotion, since emotions seem to inherit the intentional objects of the appraisals that elicit them. Anger elicited by a perception of a bully seems to be about that bully.

The enactive theory presents a principle for inclusion and exclusion: emotions, like camera modes, are states that modify action-systems to facilitate certain action-types. According to this principle, states of bodily arousal, episodic motivational changes, and social expressions are all included, and appraisals are excluded. This theory is good if it helps make sense of a broad range of emotional phenomena. In the remainder of the paper I will show how this theory helps explain the narrow but significant range of our emotional lives concerning anger and forgiveness.

4. Enactive Forgiveness

Before developing my enactive account of forgiveness I will review the facts that the theory should explain: 1) that forgiveness coincides with the end of anger, 2) that the anger in question seems to be directed at an individual for some wrongdoing (or apparent wrongdoing), 3) that we expect the wrongdoer to fix the harm that resulted from their wrongdoing, when reasonably possible, to earn forgiveness, 4) that fixing harm is neither necessary nor sufficient for earning forgiveness, 5) and that earning forgiveness seems to require establishing one's moral credibility in a forward-looking manner.

I will begin by describing a simple example where a wrongdoer harms a victim, the wrongdoer seeks forgiveness, and the victim grants forgiveness. Rowan and Harper are classmates in kindergarten. Harper brought to school a minifigure to show their friends. Rowan began to play with the minifigure without asking permission. By the time Harper noticed Rowan playing with the toy, it was already missing some of its pieces (a helmet and other accessories). Harper became very angry. Rowan admitted wrong-doing, returned the figure, and made an earnest, though unsuccessful, attempt to find the missing pieces. It became clear that the pieces were unlikely to be found any time soon. Harper, no longer angry, forgave Rowan, and they went on to play together later that afternoon.

The enactive explanation of Harper's initial anger is straightforward. Harper has been wronged. If Harper does nothing then they will risk not getting their toy back, and also send the message to others that they are a safe target for future mistreatment. Harper wants Rowan to repair the harm to the best of their ability, and credibly demonstrate an intention to treat Rowan better in the future. This will provide Harper with the benefit of repair, the benefit of improved treatment in the future from Rowan, and the benefit of discouraging future mistreatment from others. Harper may not be aware that their anger will help secure these benefits, but they explain why Harper's anger is fitting.

The most direct way for Harper to assert their rights is to physically attack Rowan until Rowan agrees to cooperate. Harper's bodily arousal and other associated bodily changes prepare them for a physical conflict, and corresponding motivational changes steer Harper towards aggressive, risky actions.¹¹ This is, however, not permitted in kindergarten, and discouraged by Harper's parents.

Fortunately, Harper's social displays allow them to assert their rights in multiple other ways. First, since Rowan can perceive Harper's readiness and willingness to fight, they will have a reason to mollify Harper. This will allow Harper to extract concessions without the risks and costs of actual physical conflict. However, being in an environment where retributive violence is not allowed, like a well-ordered kindergarten classroom, undermines Harper's ability to use the threat of violence to extract concessions.

We will now turn to a second beneficial effect of Harper's social display. Angry individuals are willing to persist and tolerate risk in physical struggles, but they will do the same in struggles that do not involve physical conflict. This could include severing the individual's relationship with a wrongdoer, and attempting to lower the wrongdoer's standing with others. Not only does anger motivate one to use these social retaliation strategies in the moment, but it also provides evidence that the victim will nurse a grudge, and persist in these efforts. The risk of the victim retaliating in this way may be sufficient to force concessions from the wrongdoer. Even a well-ordered kindergarten environment cannot entirely undermine the credibility of this threat.

¹¹ See Jennifer S. Lerner and Dacher Keltner, "Fear, Anger, and Risk," *Journal of Personality and Social Psychology* 81, no. 1 (2001).

When the story is told in this way it is no wonder why Rowan seeks forgiveness, and why Harper forgives. Anger is useful because it helps the victim extract concessions and protect their social status. The wrongdoer seeks forgiveness because the anger does its job: the credible threat of retaliation motivates them to satisfy the victim. At that point further retaliation would be useless, because the victim has already received as much compensation as they can reasonably expect. If Harper were to remain angry regardless of Rowan's compensations then there would be no benefit for Rowan to compensate Harper at all.

We can describe this dynamic using game theory. Rowan and Harper collaborate as members of their classroom community. There is always a risk that a collaborative partner will make a move that, in the short term, benefits them while harming their partner. We can diminish this risk by adopting a tit for tat strategy.¹² We return good with good, and bad with bad. The purpose of adopting this strategy is not to inflict the greatest harm on wrongdoers. Instead, we make it rational for others to treat us well. When someone has that strategy, and has the power to inflict harm on you, it is in your interest not to harm them in the first place, and failing that, to quickly redeem yourself in their eyes.

Anger and forgiveness together implement the tit for tat strategy. We need to reliably respond to bad with bad, even when it is difficult, expensive, and risky. Anger makes retaliation less difficult and expensive, it makes us more willing to tolerate the expense and risk, and it makes our readiness and willingness evident to others. We also need to be prepared to respond to good with good, even after we have been harmed in the past. This is why forgiveness is earned through credible promises of good treatment in the future.

Rowan's anger plays a productive role in this story because Harper has the power to compensate Rowan, and Rowan has leverage over Harper. If the toy was lost, and nobody was at all to blame, there would be no use mobilizing to extract concessions. If the toy was lost, and Harper was a bully that Rowan had no hope of resisting, then mobilizing for aggression would be counter-productive. The function of anger is to create or improve affordances, to enhance the agent's grip

¹² See Uzi Segal and Joel Sobel, "Tit for Tat: Foundations of Preferences for Reciprocity in Strategic Settings," *Journal of Economic Theory* 136, no. 1 (2007).

on the objects within their reach. Anger only plays that role if that grip is achievable. Otherwise it might be more suitable to respond with sadness, in the case of the blamelessly lost toy, or fear, in response to the implacable bully.

5. Explaining the Data

Now that this basic story has been presented we can see how well it does at explaining the phenomena listed above, and see which issues require further explanation.

1. Why does forgiveness coincide with the end of anger? Like all emotions, anger enhances a class of actions. Specifically, anger assists the victim in extracting concessions and credible evidence of good intentions. Forgiveness marks the end of that effort, when anger is no longer necessary or appropriate.
2. Why do we expect the wrongdoer to fix the harm, if possible, to be forgiven? First, because we cannot trust a wrongdoer who is not willing, to the extent that they are able, to fix the damage that they have done. If the victim settles for less then they are failing to apply the tit-for-tat strategy, and making themselves vulnerable to further mistreatment. Second, because fixing the harm is itself beneficial to the victim.
3. Why is fixing the harm neither necessary nor sufficient for forgiveness? It is not necessary because repair is sometimes not possible or realistically achievable. In those cases it may still be beneficial to repair the relationship. It is not sufficient because it is unwise to maintain a relationship with someone who is likely to wrong you again, even if they are sometimes willing to repair the harm that they do.
4. Why does earning forgiveness require establishing one's moral credibility in a forward-looking manner? Because that is necessary for the victim to justify resuming the relationship.
5. Why is it sometimes rational to forgive? First, because forgiveness makes it possible to resume a normal relationship, and we forgive when resuming would be beneficial. Second, because it only makes sense for the wrongdoer to comply with the demands we make in anger because we will forgive them if they accept our demands.

6. Why is it rational to seek forgiveness, even at personal cost? Because that allows the wrongdoer to convince the victim to end their hostility, and possibly resume a beneficial relationship.

There is one more critical question that I have not yet answered: why does the wrongdoer seek forgiveness in cases where the victim does not have any way to harm or benefit the wrongdoer?

I will turn this into a more general objection: my account of forgiveness presents agents as motivated primarily, or even exclusively, to further their own interests. Relationships themselves are characterized by how they benefit each participant. This is contradicted by cases where the victim is powerless to retaliate against the wrongdoer,¹³ but it also fails to do justice to the moral dimension of forgiveness, repentance, and human relationships in general. We apologize and seek to fix the harm we've done because we feel guilty, or think it is the right thing to do. We forgive because we think the wrongdoer deserves to be forgiven.

At the end of the introduction I noted that the dynamics of anger and forgiveness are not uniform across all cases, and that my strategy was to develop a simple model of those dynamics that explains the puzzling features that I identified. That simple model is now complete. The next step will be to develop more complex variants of the model.

6. Function and Motivation

Like many other people, I often check my hair when I see my reflection in the mirror. An evolutionary psychologist might say that I desire to look good because my ancestors attended to their appearances, and this enhanced their reproductive success. If this is true, it would not mean that my own self-grooming is motivated by a desire to produce many offspring. I groom myself because I am vain. The vanity of my ancestors helps explain both my existence and my own vanity, because it contributed to their reproductive success and the reproduction of their genes.

We can apply that same distinction to my analysis of anger and forgiveness. I argued that it is beneficial, in response to being wronged, to use anger to extract

¹³ Luc Bovens, „Must I Be Forgiven?,” *Analysis* 69, no. 2 (2009), 228.

compensation and reassurances of future reliability, and then forgive once we have successfully extracted as much as reasonably possible. This gives the impression that we calculate when to express anger and when to forgive. Some people do calculate in that way, at least some of the time, but we often experience anger and forgiveness as spontaneous responses.

This experience of spontaneity is still compatible with my explanation. We become angry as a result of mechanisms that we cannot directly control, and that we only influence indirectly via emotion regulation strategies.¹⁴ The same is true with forgiveness. These mechanisms are organized in such a way that they help us conform to a tit-for-tat strategy, even when we do not aim to do so. We succeed socially because our spontaneous, authentic responses conform to this strategy.

It is worth noting that emotions are blunt tools. They modify our motoric and social affordances in quite general ways, based on appraisal processes that respond to quite general situation-types. This leads to many counter-productive effects. Emotional arousal might interfere with deliberation, keep us awake at night, or damage our cardiovascular system. Social displays of anger may lead us into conflicts that we do not have the resources to resolve effectively, reveal too much about our desires and insecurities, or violate social norms. The solution to all of these problems is emotion regulation of one kind or another, which typically requires a conscious and deliberate attention to our spontaneous responses.

The more difficult question is why we seek forgiveness from individuals who will not help or harm us either way. Seeking forgiveness can be expensive. It may require costly efforts to repair harm. It would be functionally optimal, in a narrow sense, to only expend these resources when one has something to gain from the relationship.

The key to explaining these cases is that the wrongdoer needs to earn forgiveness by convincing the victim that they would not do the same thing again. The wrongdoer might form the intention to treat the victim better because they now realize that the victim has the power to either harm them or withhold significant benefits. If the wrongdoer credibly expresses this newfound appreciation for the power of the victim then that may be enough to earn forgiveness (if combined with adequate reparation).

¹⁴ See James J. Gross, "Emotion Regulation: Current Status and Future Prospects," *Psychological Inquiry* 26, no. 1 (2015).

The problem with that sort of apology is that it is contingent on the victim continuing to have the power to do harm or withhold benefits. The apology would lose its force if the victim loses their power, or if the wrongdoer changes their assessment of the victim's power. A more trustworthy apology would be based on the wrongdoer's conviction, credibly expressed, that the victim does not deserve to be wronged regardless of their power to retaliate. If the wrongdoer can credibly express this conviction then they will be easy to forgive, and being easily forgiven is a beneficial quality for a social agent. I will call this sort of person a Forgivable. A Forgivable will make expensive efforts to seek the forgiveness of those who would not have the power to extract an apology by force, because their efforts to earn forgiveness are not contingent on the victim's power. Guilt and shame play important roles in this dynamic. If someone wrongs you, that is defeasible evidence that they will not be a good social partner. However, expressions of guilt and shame indicate that the victim's anger has struck a nerve. By responding with guilt or shame, the wrongdoer demonstrates that they accept the victim's judgment that they have done wrong, and they accept that they are obligated to seek forgiveness. Failing to make a valid effort to earn forgiveness would itself lower their moral status, and justify the persistence of shame or guilt.¹⁵

Accepting moral responsibility is painful, and we are often reluctant to do it. We often prefer to hide our wrongdoing. However, someone who is unwilling to accept social sanction is someone who cannot be trusted. 'Shameless' is a derogatory term for a reason.

7. Collective Memory

There is one remaining question that I introduced in the introduction. How should we understand moral status and its relationship to anger and forgiveness? Bovens¹⁶ argues that we seek forgiveness with the goal of restoring moral status. We humble ourselves by giving the victim the power to decide whether our moral status should be restored. In my analysis from the previous section I argue that a

¹⁵ It is possible that the victim will be unreasonable in refusing to forgive, which is why making a valid effort to earn forgiveness is sufficient.

¹⁶ Bovens, „Must I Be Forgiven?,” 230.

Forgivable is an individual who is motivated to seek forgiveness regardless of the victim's power. They see the pursuit of forgiveness as essential for restoring their moral status.

This leaves open the question of what constitutes our moral status, and how we can give someone the power to restore it. I am going to define moral status, within this analysis, as the sum of what makes us feel moral shame or moral pride. My proposal is that this moral status is constituted, at least in part, by all of the moral judgments that are made about us by everyone that matters to us. It is how we appear in the collective memory of our community. If someone remembers an unkind thing that I said, that goes on my record. If someone remembers me acting generously, that does too.

There are two points to clarify about the contents of this memory. First, some people's moral judgments have more weight than others for us. We tend to care more about our self-evaluations, and judgments about us from close friends, family, and certain authority figures.

Second, these judgments count even when we do not accept their validity. It may seem like I am confusing reputation and virtue. However, while reputation and virtue are obviously distinct, we typically care about both. I am defining moral status by what has the power to produce shame or pride, and as Calhoun¹⁷ points out, everyone we care about has the power to shame us, even when we reject the judgments they make. It is not psychologically realistic to only be morally shamed when we agree with those judgments.

This has both good and bad consequences. On the negative side, we can suffer shame unjustly, based on misperceptions or prejudices. On the plus side, this dynamic makes us accountable to each other. Being accountable in this way makes us dependent on the good will of others, and vulnerable to their irrational grudges. However, by giving this power to the people we care about, we make ourselves more trustworthy, and we become a Forgivable.

We can now integrate collective memory into our account of the relationship between anger and forgiveness. Anger is a mobilization of mind and body that facilitates aggressive action. Its function is to enhance affordances for certain types

¹⁷ Cheshire Calhoun, "An Apology for Moral Shame," *Journal of Political Philosophy* 12, no. 2 (2004): 138.

of action, rather than to represent a core-relational theme, so it is fitting whenever it is appropriate to prepare for those actions.

One use of anger is to extract concessions from those who have harmed us. These concessions are typically getting the wrongdoer to repair the harm, to the extent that it is reasonably possible, and to convince the victim that they will be better in the future. Anger is effective at extracting concessions because 1) it prepares the victim for a physical struggle, 2) social displays of anger provide the wrongdoer with evidence that the victim is prepared for a physical struggle, and 3) social displays of anger also provide evidence that the victim is ready for a social struggle.

With the concept of collective memory, we can introduce a fourth way in which anger can help extract concessions. Expressions of anger are reliable sources of evidence that negative information about the wrongdoer has entered the collective memory about them. This will impact the wrongdoer's moral status, and the extent of this impact will depend on the importance of the victim to the wrongdoer, whether other individuals share that assessment, and the degree to which the wrongdoer shares that assessment of themselves.

Anger is effective at extracting concessions in part because it normally ends when sufficient concessions are delivered. Along with the end of anger, forgiveness also brings an amendment to the victim's portion of the collective memory. This revision does not erase the record of the wrongdoing, but the wrongdoer is no longer marked with a moral demerit for that action. The wrongdoer, knowing that they have been forgiven, is reassured that the victim will no longer resent them for their past behavior, and that they no longer deserve to be resented.

8. Deliberate Forgiveness

Up to this point I have focused on cases of spontaneous anger and spontaneous forgiveness. I started with a puzzle about the temporality of anger: that anger is typically conceived of as a representation of a past act of wrongdoing, yet our anger ends with forgiveness based on the wrongdoer's credible promise of good behavior in the future. I resolved this discrepancy by providing a future-oriented

psychological account of anger. Forgiveness occurs when the wrongdoer's credible promise produces a spontaneous end to the victim's anger, which on my account is fitting. However, we sometimes choose to forgive, and deliberate about whether to forgive. I will develop an account of these cases that is based on emotion regulation.

I'll introduce my account using Brandon Warmke's¹⁸ discussion of the normativity of forgiveness. He cites Strawson's¹⁹ observation that the victim may no longer blame the wrongdoer for wrongs that have been forgiven, and the wrongdoer is no longer obligated to make amends to the victim. Warke's aim is to explain how forgiveness alters the norms that govern both the victim and wrongdoer.

After considering a number of possible explanations for the normative effects of forgiveness, Warmke introduces the view that forgiveness is a declarative act, in which "Forgivers alter the relevant norms of interaction between victim and wrongdoer by declaring the wrongdoer forgiven."²⁰ Warmke argues that this is stronger than Pettigrove's²¹ view that forgiveness is a form of commitment, since commitments can be withdrawn, but it is not permissible to withdraw one's forgiveness and resume one's resentment. I will not attempt to resolve whether the declarative or commissive theory is stronger. Instead, I will use my own theory to explain Warmke's normative data.

My analysis of spontaneous anger explains why forgiveness creates these normative changes. Anger ends with forgiveness when the victim is satisfied by the wrongdoer's concessions. In this manner the victim implements the tit for tat strategy. Anger is a form of aggression. The wrongdoer responds to that aggression with concessions because the wrongdoer trusts the victim to stick with a tit for tat strategy and return good with good. An individual who returns to anger after giving forgiveness cannot be trusted, and society informally sanctions those who do. It is therefore beneficial for the victim to be the sort of person whose anger ends, and remains gone, after they have been adequately compensated.

¹⁸ See Brandon Warmke, "The Normative Significance of Forgiveness," *Australasian Journal of Philosophy* 94, no. 4 (2016).

¹⁹ See Peter F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962).

²⁰ Warmke, "The Normative Significance of Forgiveness," 12.

²¹ See Glen Pettigrove, *Forgiveness and Love* (Oxford: Oxford University Press, 2012).

Our spontaneous anger sometimes conforms to these norms, but not always. We might remain angry when it is time to forgive, become angry again after we have forgiven, or even forgive when we should not. It is therefore beneficial to deliberate about when our anger should end, and to make sure that we do not resume aggression after we have publicly forgiven our wrongdoer. Emotional reactions are blunt tools, but emotion-regulation allows us to sharpen their social effects.²²

I observed earlier that there is a difference between a functional explanation for a psychological feature and the agent's own reasons. Natural selection explains my vanity, but I don't look in the mirror in order to reproduce. The same is true for emotion-regulation. I deliberate over whether to forgive, I use emotion-regulation strategies to conform to norms governing anger and forgiveness, and I participate in sanctioning individuals who violate those norms. This allows me to maintain my status as an individual who is known to implement a tit for tat strategy. However, I do not do this because I desire to be known in that way. Instead, I have internalized a broad range of norms governing inter-personal relations, and these include norms governing anger and forgiveness. I follow these norms because they are authentically mine.

This theory is therefore an explanation of two related facts: a fact about the temporality of anger and a fact about the norms governing forgiveness. Forgiveness depends on forward-looking criteria because anger is an enactive process that implements a tit for tat social dynamic. We have norms of forgiveness that do not permit us to renew our aggression after forgiving because internalizing those norms, and using emotion-regulation to conform to them, allows us to implement that tit for tat dynamic with greater precision.

²² This is an example of deliberately instituting a social dynamic that is in other cases produced by spontaneous emotional responses. I have argued elsewhere that fake laughter often plays a similar role. The emotion of amusement makes an individual more tolerant of harmless norm violation, and through emotional contagion creates a social dynamic that is conducive to informal bonding. Fake laughter serves as an invitation to collectively create the same dynamic, even in the absence of actual amusement. Daniel Shargel, "What We Can Learn from Forced Laughter," in *The Moral Psychology of Amusement*, ed. Brian Robinson (Lanham: Rowman & Littlefield, 2021), 13.

9. Conclusion

The enactive theory of emotion helps us understand some puzzling features of the dynamics of anger and forgiveness. Typically anger is understood as backwards-looking. It is fitting when it accurately represents past events. However, we end our anger, and expect others to end their anger, based on forward-looking grounds for forgiveness. It is certainly possible to develop an explanation of how forward-looking criteria for forgiveness are compatible with a backwards-looking view of anger. There is always a bullet to bite. However, I have presented a forward-looking theory of anger that provides a simple and versatile explanation for the various forms of forgiveness that we find in the wild.

On the enactive theory that I have defended, emotions temporarily modify various physiological and psychological features to facilitate certain types of actions. In the simplest model of forgiveness the victim prepares to retaliate via physical or social means, and the threat of retaliation compels the wrongdoer to satisfy the victim's demands. Once the victim is satisfied it is no longer necessary or beneficial for them to mobilize for retaliation, so their anger ends. In a more complex model, the anger of the victim induces the wrongdoer to perceive that their moral status has been lowered, and this motivates them to redeem themselves. The logic of these two cases is the same from the perspective of the victim. Their anger has successfully reshaped their social world, and once it has accomplished its goal it ends.

There are endless further variations, due to the varieties of possible social relations and patterns of cognition. A victim may, inspired by theology, offer unconditional forgiveness, or with the guidance of a therapist pursue cognitive behavioral therapy to address persistent resentments. A wrongdoer may see their partner's anger, and respond with care inspired by love rather than guilt or shame. These are all compatible with an understanding of anger as a mobilization of the body and mind for aggressive action, and forgiveness as a demobilization of that same response.

Bibliography

- Bovens, Luc. "Must I Be Forgiven?" *Analysis* 69, no. 2 (2009): 227–233.
- Calhoun, Cheshire. "An Apology for Moral Shame." *Journal of Political Philosophy* 12, no. 2 (2004): 127–149.
- D'Arms, Justin, and Daniel Jacobson. "The Moralistic Fallacy: On the Appropriateness of Emotions." *Philosophical and Phenomenological Research* 61, no. 1 (2000): 65–90.
- Deonna, Julien, and Fabrice Teroni. *The Emotions: A Philosophical Introduction*. London: Routledge, 2012.
- Gibson, James J. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin, 1979.
- Gross, James J. "Emotion Regulation: Current Status and Future Prospects." *Psychological Inquiry* 26, no. 1 (2015): 1–26.
- Kenny, Anthony. *Action, Emotion, and Will*. New York: Humanities Press, 1963.
- Lerner, Jennifer S., and Dacher Keltner. "Fear, Anger, and Risk." *Journal of Personality and Social Psychology* 81, no. 1 (2001): 146–159.
- Pettigrove, Glen. *Forgiveness and Love*. Oxford: Oxford University Press, 2012.
- Russell, James A. "Core Affect and the Psychological Construction of Emotion." *Psychological Review* 110, no. 1 (2003): 145–172.
- Scherer, Klaus R. "The Dynamic Architecture of Emotion: Evidence for the Component Process Model." *Cognition and Emotion* 23, no. 7 (2009): 1307–1351.
- Searle, John R. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press, 1983.
- Segal, Uzi, and Joel Sobel. "Tit for Tat: Foundations of Preferences for Reciprocity in Strategic Settings." *Journal of Economic Theory* 136, no. 1 (2007): 197–216.
- Shargel, Daniel. "What We Can Learn from Forced Laughter." In *The Moral Psychology of Amusement*, edited by Brian Robinson, 13–30. Lanham: Rowman & Littlefield, 2021.
- Shargel, Daniel, and Jesse J. Prinz. "An Enactivist Theory of Emotional Content." In *The Ontology of Emotions*, edited by Hichem Naar and Fabrice Teroni, 110–129. London: Routledge, 2018.
- Strawson, Peter F. "Freedom and Resentment." *Proceedings of the British Academy* 48 (1962): 1–25.
- Varela, Francisco J. "Whence the Origin of Perception? A Cartography of Current Ideas." In *Understanding Origins: Contemporary Ideas on the Origin of Life, Mind and Society*, edited by Francisco J. Varela and Jean-Pierre Dupuy, 235–265. Boston: Kluwer, 1991.
- Warmke, Brandon. "The Normative Significance of Forgiveness." *Australasian Journal of Philosophy* 94, no. 4 (2016): 687–703.

Streszczenie

Enaktywistyczna koncepcja gniewu i przebaczenia

Zdaniem wielu teoretyków, gniew stanowi reakcję na akt wyrządzenia krzywdy. Przebaczenie kładzie mu kres i często uważa się, że sprawca może na nie zasłużyć poprzez naprawienie szkody wyrządzonej ofierze. Trudno jednak pogodzić takie ujęcie gniewu z następującymi faktami: skuteczne naprawienie szkody nie jest ani warunkiem koniecznym, ani wystarczającym uzyskania przebaczenia, natomiast konieczne jest, aby sprawca wykazał, że w przyszłości będzie godny zaufania. W artykule przedstawiam wyjaśnienie relacji pomiędzy gniewem a przebaczeniem oparte na enaktywistycznej teorii emocji. Zgodnie z tym ujęciem, emocja jest adekwatna, jeśli przygotowuje podmiot do skutecznego wykonania określonej klasy działań. Gniew mobilizuje umysł i ciało ofiary do uzyskania ustępstw ze strony sprawcy. Przebaczenie wyznacza moment, w którym przygotowywanie się do agresywnego działania wobec sprawcy przestaje być reakcją adekwatną.

Słowa kluczowe: emocje, afordancje, gniew, przebaczenie, psychologia moralna

Zusammenfassung

Ein enaktivistischer Ansatz zu Wut und Vergebung

Nach Ansicht vieler Theoretiker ist Wut eine Reaktion auf eine verletzende Handlung. Vergebung beendet die Wut, und oft wird angenommen, dass der Täter sie sich verdienen kann, indem er den dem Opfer zugefügten Schaden wiedergutmacht. Es ist jedoch schwierig, eine solche Sichtweise auf Wut mit den folgenden Tatsachen in Einklang zu bringen: Eine wirksame Wiedergutmachung des Schadens ist weder eine notwendige noch eine hinreichende Bedingung für die Gewährung von Vergebung; vielmehr ist es erforderlich, dass der Täter zeigt, dass er in Zukunft vertrauenswürdig sein wird. In diesem Artikel stelle ich eine Erklärung der Beziehung zwischen Wut und Vergebung vor, die auf der enaktivistischen Theorie der Emotionen basiert. Gemäß diesem Ansatz ist eine Emotion adäquat, wenn sie das Subjekt auf die erfolgreiche Ausführung einer bestimmten Klasse von Handlungen vorbereitet. Wut mobilisiert Geist und Körper des Opfers, um Zugeständnisse seitens des Täters zu erwirken. Vergebung markiert den Moment, in dem die Vorbereitung auf aggressive Handlungen gegenüber dem Täter keine adäquate Reaktion mehr darstellt.

Schlüsselworte: Emotionen, Affordanz, Wut, Vergebung, Moralpsychologie

Ins Deutsche übersetzt von Anna Pastuszka

